ESTIMATION OF CROP YIELDS

43464

GRICULTURE ORGANIZATION OF THE UNITED NATIONS

ROME 1954

YEARBOOK OF FOOD AND AGRICULTURAL STATISTICS 1953

PART I PRODUCTION

Part I provides basic statistics on world agricultural production. It gives authoritative information on land use, agricultural population, crops, livestock numbers and products, food supplies and their utilization, and on commercial fertilizers, pesticides, and agricultural machinery. It also includes the more important series of agricultural commodity prices in many countries, as well as index numbers of prices received and paid by farmers, and of agricultural production.

PART II TRADE

The second part is a basic reference work on world trade in agricultural products and gives statistics of the imports and exports of the major agricultural commodities. It includes regional and world totals, computed from official and unofficial information. For some major commodities, data are given by trade season as well as by calendar year.

This two-volume yearbook is a valuable reference work for importers, exporters, commercial houses, as well as official bodies concerned with food and agriculture. It is available from any book-seller, or from any of the FAO sales agents listed on the back cover of this publication.

Price: Each part, $3.50 or 17s. 6d.

P 196 E

ESTIMATION OF CROP YIELDS

Errata

On page 51, line 29 read:

This will be the third corner of the plot provided the distance of the diagonal between the first peg and this point is 36' $10\frac{1}{2}$". Check the measurement of the diagonal over the ear-heads.

On page 52, line 8 read: field, not fidle.

On page 52, line 21 read:

It is advisable not to allow the surrounding crop of the field to be harvested until the crop within the plot is harvested and removed to the threshing ground.

On page 52, line 27 read: hour, not hojr.

ESTIMATION OF CROP YIELDS

by

V.G. PANSE

Consultant, Statistics Branch,
Economics Division

43464

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

Rome, Italy July, 1954

Printed in Italy

CONTENTS

ESTIMATION OF CROP YIELDS

FOREWORD

Among the major tasks of the Economics Division of the Food and Agriculture Organization is the promotion of the improvement of national systems for collecting and reporting agricultural statistics. Statistical information is fundamental for setting up plans and taking action for developing agricultural production, for improving the distribution of food and other products, and for raising the general standard of living of populations.

The program of work of FAO, as approved by the Conference in 1952, included therefore the preparation and distribution of a manual on the theory of sample surveys with illustrative examples from the field of agricultural statistics. This has been done (1). A number of manuals on specific applications were also to be prepared in order to place at the disposal of member countries sufficient details, in an analytical presentation, on modern statistical methods used in the more advanced countries. The present manual on estimation of crop yields is the first of the series. It has been prepared not only as a follow-up of the general manual on sample surveys but also to meet the recommendation formulated by the First Conference of Statisticians of Countries in Africa South of the Sahara, which met in Salisbury, Southern Rhodesia, in 1951. That recommendation was expressed as follows: "The Conference feels that the problem of measuring yields together with the problem of measuring areas should be brought to the attention of FAO with a view to the publication of a survey of the work which is being done currently throughout the world. The Conference wishes to lay on record its opinion that a critical survey of crop-cutting methods is an urgent priority if the needs of territories such as Africa South of the Sahara are to be met."

The Food and Agriculture Organization was fortunate to have secured Dr. V.G. Panse's services to prepare this manual. Dr. Panse is well known for his achievements in India in making use of probability

---

(1) P.V. Sukhatme: "Sampling Theory of Surveys with Applications", published by the Iowa State College Press, Ames, Iowa, U.S.A., and the Indian Society of Agricultural Statistics, New Delhi.

sampling methods in the field of agricultural research. His work and association with the Indian Council of Agricultural Research have largely contributed to make the crop-cutting method one which can now be considered as the most reliable of current methods for estimating crop yields.

In the present manual he gives a short account of current methods and a full analysis of crop-cutting procedures, with technical considerations and practical explanations on field work. The numerous examples drawn from various parts of the world show how the method works and how it should be adapted under different crop conditions. It is our hope that this manual will meet the need expressed for more information on this technical field and that it will raise interest in the application of objective methods for improving the accuracy of our knowledge of crop yields.

A. H. Boerma
Director, Economics Division

I. INTRODUCTION

Reliable estimates of annual production of food crops and other agricultural commodities are assuming a rapidly growing importance, as countries make serious efforts to tackle the problem of feeding their populations and of otherwise planning national economies to raise their standard of living. Direct measurement of total production of a crop with any degree of reliability is, however, possible only for those crops which are either marketed entirely or of which the cultivation is confined to a few large estates. Examples are jute, tea, cocoa, rubber. For such crops accurate data can be procured through market organizations dealing with the particular commodities or from the farms or estates which produce them. For the large majority of crops including food crops, however, which are grown by millions of farmers in scattered holdings, anything like a complete enumeration for measuring their annual production is impossible. The practicable approach to the measurement of production of such crops is to consider the two components of total production of a crop, namely its area and yield rate, separately, and measure or estimate each; the product of the two then gives the total output of the crop. In the present publication we are concerned only with the problem of estimating the yield rate. A special aspect of this problem is forecasting yield, i.e. obtaining reliable pre-harvest indications of what the harvested yield is going to be. Since physical measurement of yield is ordinarily not possible at the stage when yield forecasts are made, the latter have necessarily to depend upon the knowledge of factors correlated with yield, such as the condition of the crops as judged by the personal judgment of the observer, the measurement of plant characters and previous and current meteorological conditions. We shall, however, not deal with the subject of forecasting crop yield, but confine our attention to the measurement of yield at harvest.

II. CURRENT METHODS OF ESTIMATING CROP YIELDS

Statistics of crop yields in most countries are based on periodic reports from crop reporters. Essentially such reports represent the reporters' quantitative judgment of what the yield is, based on their personal impression of the crop and on enquiry among farmers. In more advanced countries such as the U.S.A. and several European countries, crop reporters are farmers and other private individuals resident in rural areas and connected with farming, who voluntarily supply to the government agency concerned, the information called for in a mailed questionnaire. In underdeveloped countries, crop reporters are generally government officials or agents who submit reports relating to crop yields for the area under their administrative charge. These official reporters are naturally much less numerous than voluntary crop reporters in advanced countries, and consequently, the unit or area for which a report is made is usually large, such as an administrative sub-division of a district and possibly a village, but hardly ever a cultivator's holding or a field. Estimates of yield derived from such reports are likely to be much more inaccurate than those for which a very large number of reporters supply data on their respective areas.

Unfortunately, the most serious defect in yield estimates based on crop reporters' data, whether the reporters are official or voluntary, is that they are subject to large and indeterminate biases. Investigations made in India have shown that the official reporters have a marked tendency to lean towards the normal, with the result that yield is underestimated in favourable seasons and overestimated in poor seasons. Voluntary crop reporters, on the other hand, might have incentives for a systematic underestimation of yield. In the U.S.A., the regression method is used (1)* as a regular routine for eliminating

* *Numbers in brackets correspond to references given at the end*

bias in the crop reporters' estimates, the regression of these estimates on "actual" yields for past years being used for this purpose. The "actual" yields are provided by the revised estimates of yield late in the season, on the basis of check data from various official as well as private sources, such as marketings, shipments, the amount of the crop processed or handled in factories, and other reasonably complete utilization information. Such data are available for a few crops like cotton, tobacco, sugar beet; but even here there can be no guarantee that the information is really complete. For the majority of crops, including food crops which are partly consumed locally or are stored by the farmer for his own use and for feeding his livestock, such check data are not available, and indications of yield from other sources, such as an annual farm census in a few states or the quinquennial census in all states, are used as a basis for revising crop reporters' estimates of yield. It is possible, however, that the data from these other sources also suffer from biases as the crop reporters' estimates do. Under the conditions prevailing in most countries, it is not possible to obtain worthwhile information on the total production of the majority of crops such as would permit the systematic application of the regression technique in order to improve the quality of the crop reporters' estimates of yield. Further, where such regression is feasible, it can remove only the bias for which a trend has been established. Biases resulting from the influence of special conditions, such as for example floods, drought or other extremely unfavourable weather situation, or a very favourable season resulting in a bumper crop, or sudden shortages in supply, cannot be dealt with effectively by this method. As an example, the estimate of cotton production in the U.S.A. in 1951 was found to be 14 percent incorrect. Experience in other advanced countries is similar. It has been shown that during the war years, production of bread grains was underestimated in Sweden by as much as 20 percent and, what was even more serious, the tendency to underestimate was greater in bad years (2). In Germany, owing to the imposition of economic control measures during the war and postwar years, and under mounting pressure on the farmers to increase deliveries, crop reporters have tended to underestimate

yields rather consistently, so that the demand upon the farmers in their districts could be kept down to what they considered a reasonable level (3). No accurate measurement of the degree of understatement was available. The Ministry of Agriculture's estimates of the yield per acre of the potato crop in England have long appeared to be very low, and this was confirmed by sampling surveys projected for objective estimation of yield since 1948 (4). In India, with the control of the sale and distribution of food grains introduced during the war years, the official estimates of production of food crops have been found to be 6 to 7 percent lower on the average when compared with the results of crop sampling surveys, underestimation in individual states being sometimes much higher. Experience during the recent agricultural census in Southern Rhodesia indicated that the difficulties of using eye estimates of yield are infinitely greater under African conditions than they are in Europe, particularly in areas where mixed seed is sown. The conclusion drawn was that the eventual solution will probably be to abandon eye-estimation completely, and to use instead the measured output of a random selection of plots in a sample of villages (5). Thus, experience under a wide range of agricultural and economic conditions demonstrates the inability of subjective or personal methods of estimating yield to provide reliable results and points to the necessity of replacing such methods by those involving:

(i) the selection of a representative sample of the crop for observation; and,

(ii) using for observation the method of physical measurement of yield at harvest.

Eye estimation of yield in place of physical measurement on a representative sample crop has sometimes been suggested as adequate for improving the accuracy of yield estimates. The chief defect with eye estimates of yield would, however, appear to be a lack of consistent and close correlation with actual yield due to unpredictable biases of the observers, and such a method is not likely to result in any appreciable improvement.

If accurate and unbiased yield estimates are required for such purposes as planning agricultural production and formulating material food policies, there is no alternative but to resort to objective methods of estimation on a random sampling basis.

III. THE TECHNIQUE OF OBJECTIVE MEASUREMENT OF YIELD BY CROP CUTTING

Estimation of yield by crop cutting involves the selection of a representative sample of a crop over the area which is being studied, and harvesting and weighing the produce from each of the several units constituting the sample. The sample units actually harvested are plots of prescribed dimensions located and marked according to clearly defined procedures. These plots may be located in a single stage of random selection over the entire area of the crop, or more commonly by first selecting villages or farms at random out of all villages or farms in the area which grow the crop, then selecting at the next stage one or more random fields growing the crop within these primary units, and finally marking in one or more random positions in the selected fields the sample plots for harvesting. That this procedure is practicable and capable of giving yield estimates free from bias and possessing a high degree of accuracy, has been demonstrated by several large-scale crop cutting surveys that have been carried out in different countries.

India

India has perhaps accumulated the largest experience in this field over the last 12 or 13 years (6). Under the leadership of the Indian Council of Agricultural Research, investigations carried out first on cotton and then on rice and wheat have shown that the crop cutting method can be employed annually for estimating yield through the administrative machinery of the states. With this demonstration, the states have taken up crop cutting surveys on a rapidly expanding scale, so that by 1952, 75 percent of the area under various food grains was being sampled by this method. The percentage of the area under wheat and rice crops that was sampled was still higher. Investigations for extending the method to crops like sugar cane, coconuts and fruits are in progress. The method essentially consists in stratifying a district, which is the administrative unit for which yield

estimates are required, by its administrative sub-divisions, selecting randomly within each sub-division a certain number of villages roughly in proportion to the area under the crop in that sub-division, and locating within each village two or three random fields under the crop for marking a crop cutting plot in a random position in each. The crop cutting plot is usually rectangular and of 50x25 (links)^2 equivalent to 1/80 acre (33 feet x $16\frac{1}{2}$ feet) in size. A certain variation in size is however permitted, depending upon the crop and the tract. In Orissa where fields are very small, the crop cutting plot is only 1/160 acre, while in cotton yield surveys the plot is as large as 1/10 acre. Whatever the size, the plot is marked with the help of pegs, string, measuring tape and a cross-staff. The procedure adopted for harvesting and processing the produce of the plot is similar to that prevalent among the farmers of the locality. The sample, therefore, provides an unbiased estimate of the produce on the harvesting date. Since the grain at harvest contains moisture, a sample of it is stored and re-weighed after drying, so that allowance for the moisture content may be made in the estimated yield. The results of a large number of surveys carried out by this method have shown that the sampling error for the district estimate of yield for principal crop growing districts, is generally 4 to 8 percent, and that for the state estimates of yield, 1 to 3 percent. On an average, between 60 to 200 crop cutting plots are harvested in each district. The ratio of the area actually harvested to the total area under the crop, often called the overall sampling fraction, varies from state to state within a range of 5 to 20 parts per million. The field work is done by the staff of the Land Records and Agricultural Departments of the State as part of their normal duties under the guidance of the State Statistician, the co-ordination of the plans and results of the surveys in different states being effected by the statistical staff at the centre. The centre also maintains a nucleus staff to exercise independent supervision over field work in the states.

Special investigations, both experimental and theoretical, have at all times formed part of this work.

These investigations have aimed at improving the overall efficiency of the surveys and are related to their various aspects, such as efficiency of stratification and of the different sizes of sample units, optimum allocation of sample units, optimum methods of selecting the sample, enumerators' biases, and methods of evaluating and reducing them, and so on. The results of these investigations have made a substantial contribution to the sampling theory of surveys and have now been incorporated in a book of that title by Sukhatme. (7)

The Indian Statistical Institute of Calcutta has also been conducting crop cutting surveys on rice and jute in Bengal State for several years (8,9). Apart from stratification by administrative sub-divisions, artificial strata marked by drawing lines parallel to the lines of latitude and longitude on the map, have been tried. Crop cutting plots have been marked in a single stage within the strata by locating points on the map with the help of random co-ordinates. Multi-stage sampling has also been tried by dividing each stratum into rectangular or square cells, selecting a certain number of these randomly and within each locating the requisite number of sample plots for crop cutting. The plot size employed has been small, marked in the earlier stages with a triangular frame which enclosed an area of 12.5 square feet. This size has been increased gradually and the standard technique adopted at present involves the use of an implement having a folding arm opening out to various lengths, e.g. 2 ft., 4 ft., 5 ft. 7½ in., and 8 ft., with which three or four concentric circular cuts are made. The field work is entrusted to ad hoc enumerators who move rapidly during the harvesting season and take sample cuts in fields which are ready for harvest at the time of their visit. The possibility of bias arising from various sources in this procedure is under investigation. The accuracy of the state estimate is high, a standard error of 1 to 2 percent being expected from about 2,500 to 3,000 sample plots.

Ceylon

On the initiative of the Food and Agriculture

Organization and under its Expanded Technical Assistance Program, the experience gained in India has been utilized for extending the crop cutting survey method for the estimation of crop yields to other countries where conditions are similar. Thus, in Ceylon an island-wide sample survey was successfully carried out in 1951/52 for estimating the yield rate for paddy (10). Sub-divisions of districts (divisional revenue officers' divisions) were used as strata. The plot size was 1/80 acre (33 feet by 16 feet 6 inches) but in terraced areas where fields are very small, the entire field was harvested as a sample unit. The selection of the crop cutting plot was made in four stages, namely those of villages, holdings, fields and crop cutting plots. The number of plots allotted to each stratum was roughly proportional to its paddy area. The produce was harvested and threshed on the same day and a sub-sample of the paddy grain was taken to estimate moisture content. A total of about 2,000 crop cutting plots were harvested at the rate of three plots per village, and gave the overall estimate of yield per acre with a standard error of 1.3 percent. For individual districts, the standard errors ranged from 4 to 8 percent. These surveys are being continued.

Thailand

The possibility of estimating the yield of paddy in Thailand by means of crop cutting surveys was demonstrated by carrying out in 1952 a random sampling survey in one province, in the course of an international training center for agricultural sampling conducted by FAO in that country (11). The list of villages in each commune, which is the administrative sub-division of a district, is the only frame available for sampling in Thailand, and as the land is not cadastrally surveyed, sketch maps had to be prepared of the selected villages for the location of crop cutting plots. Three random paddy growing fields were selected in each village with the help of this sketch map and a sample plot of 5 meters by 10 meters was harvested in each field. The plot was marked by means of pegs, string, measuring tape and a cross-staff. The results indicated that if a sample of about 1,000 crop cutting plots at the

rate of 3 per village was distributed in proportion to the rice area in the different districts, it would provide an estimate of yield with a sampling error of 1 percent.

Indonesia

Sample surveys on similar lines are in progress in Indonesia under the Expanded Technical Assistance Program. The plot size of 5 meters by 10 meters used in Thailand has been found suitable.

Japan

The sampling method with objective measurements for estimating crop yields was introduced in Japan in 1947 and is being applied as normal annual routine for the major food crops.

The territory of Japan is divided into approximately 3 million units (1 unit = 2 hectares) of land. These constitute the primary units of sampling and are grouped into several strata according to geographical conditions, and also according to high, medium and low yielding regions. The method of sampling primary units within each stratum is simple random.

For estimating yield two hitsus of land (a unit of land ownership having an official number and an average area of 650 square meters) are sub-sampled within each primary unit. Further, two plots of equal size are selected at random within each sampled hitsu, and the crop therein is harvested, dried and weighed. The plot for rice is circular in shape and about 3 square meters in size. The plots used for estimating wheat and barley are rectangular and about 1.5 square meters in size.

The above process of stratification and selecting the sample is carried out within each administrative unit for which an estimate is required. In the case of rice, for example, estimates are required for the area covered by each Branch Office of the Statistics and Survey Division

of the Ministry of Agriculture and Forestry (on an average each Branch Office is a group of five towns and villages) and the intensity of sampling is determined accordingly. In the case of wheat and barley, estimates are required for each county. For other crops estimates are needed only for each prefecture. The standard error of the estimate for each prefecture is about 0 8 percent in the case of rice and 1.5 percent in the case of wheat and barley.

Investigations carried out by Japanese Statisticians, working with very small plots ranging from 0.3 square meters to 3 square meters, show that plot sizes chosen for the purpose of the yield surveys are about the optimum in the sense that the coefficient of variation does not tend to decrease by further increasing the plot size. The question of possible biases connected with small plots, however, does not appear to have been investigated.

Basutoland

An interesting example of the application of the crop cutting method to African conditions is provided by the yield survey conducted as part of the agricultural census carried out in 1949/50 in Basutoland (12), in connection with the world census of agriculture initiated by FAO. The Basuto peasant is incapable of furnishing information on yield, and officials of the Agriculture Department also find great difficulty in making reliable estimates, owing to great variations in soil fertility and in the level of agricultural husbandry in different parts of the territory. Further, data on the disposal of the produce, such as the proportion of the crop sold and retained by the producer are insufficient, and cannot be used to estimate total production. It was, therefore, decided to adopt crop cutting for direct estimation of yield.

Maize, sorghum, wheat and peas were the principal crops, and barley, oats and beans were the minor crops that were sampled. The design adopted was a two stage sample with ES units listed during the 1946 population census as the first stage units, and fields as the second stage

units. It was decided to select twelve fields for crop cutting on each dominant crop and six fields for each minor crop. The actual number of fields sampled was determined by the speed of harvesting in progress, the possibility or otherwise of visiting late maturing fields again and other similar factors. Within each selected field, two random sample plots, each measuring 7 yards by 7 yards, or approximately 1/100 acre were located with the help of pegs and rope, the crop enclosed by the rope boundary being reaped and weighed by spring balances on the spot. The entire produce of each plot was weighed. Thus with maize and sorghum, cobs and heads were weighed; with other crops the entire plants as cut by sickle. The ratio of the weights of cobs, heads or bundle of plants to actual grain weight was determined by members of the survey teams attending the normal threshing operations of the peasants and taking samples. Similarly, the moisture content of the grain at harvest was determined from samples taken at government experimental plots. These threshing and drying factors were applied in order to calculate from the total yield of crop cutting plots, the estimates of yield per acre of storable or saleable produce. For maize, which is the most widely grown crop in the territory, the yield per acre of grain was estimated with a standard error of 3.5 percent. Crop estimation is regarded as one of the principal functions of the Agriculture Department and it was concluded that it is possible to carry it out efficiently by crop cutting as part of the normal routine of the district staff, without dislocating or adding unduly to the staff's present duties. It was suggested that while sample plots used in the present survey were satisfactory, the size should be increased to 1/50 acre. Square plots as used in the survey were also found easier to handle in the field than circles.

U.S.A.

Sample surveys for objective measurement of yield have been made in the U.S.A. also. The most extensive series reported are those on wheat (13,14). Beginning with a preliminary investigation in the eastern half of North Dakota in 1938, the surveys were extended to the States of Kansas, Oklahoma, Nebraska and South and North Dakota

in 1939 and were continued in Kansas and Oklahoma in 1940 also. From the operational stand-point, selection of a random sample of fields was not considered practical and the method of route sampling was employed instead. Each state was stratified according to crop reporting districts formed by the Agricultural Marketing Service for the purpose of crop estimation by grouping neighbouring counties together, and the roads travelled formed a gridlike pattern over each district. The geographic miles travelled in each district along these roads were approximately proportional to the total geographical area of the district. The frontage of wheat along the road was recorded on the crop-meter attached to the car. Depending upon the density of wheat acreage in the several counties, crop cutting samples were taken in fields at intervals of $\frac{1}{2}$ to 4 miles of wheat frontage along the route. Sampling was confined to a strip lying within about 100 yards from the edge of the field. Two samples were taken in each field within this strip with the help of random numbers.

Sampling was done by thrusting the two open arms of a U-shaped rigid steel frame across the drill rows. The length of each open arm of the frame was 24 inches and the distance between the two was 26.136 inches, so that the frame included an area of 1/10,000 acre. With a row spacing of 6 inches, four contiguous rows of 26.136 inches length would be included in a sample unit. Due to varied widths of drill rows found in commercial fields, however, the sample size was not fixed and the number of rows included in the sample was determined on the basis of the row spacing observed in the selected field, according to a schedule prepared for this purpose. Where the row spacing was as wide as 14 inches, only two rows were included in the sample. The yield rate for each sample unit was calculated separately by applying the conversion factor appropriate to the size of the sampling unit used. The ear heads clipped from the sample rows were packed and transported to a central laboratory where the grain was threshed, cleaned and weighed.

A total of 1,300 and 1,100 samples were taken

in Kansas and North Dakota in 1939, and 2,200 and 1,200 in Kansas and Oklahoma in 1940. The standard errors for mean yield per acre for these samples were 2 to 2.5 percent. The yields estimated from the sample would not reflect harvesting and processing losses encountered in commercial wheat production. The possibility of the estimates being biased owing to other causes also, such as : (i) the use of a very small sampling unit; (ii) the possibility that sampling routes were selective and yield levels along the routes were different from those in the area being sampled; and (iii) the restriction of the sample to a 100 yard strip of the field near the roadside and from other factors, was recognized, and sampling studies to determine the extent of bias in the present method of sampling have been suggested.

Germany

An instructive series of surveys for estimating the yield of wheat, rye and potatoes by crop cutting was started in 1948 in the British and American zone of Germany (3). The object was to obtain accurate estimates of yield of these crops for each province *(Länder)*. For this reason, the proportional allocation of the total sample over all the six provinces, according to the total acreage of the crop in each province, was modified in order to sample in any one province not fewer than 100 fields, which was considered necessary for the desired degree of accuracy of the provincial estimate, and not more than 400 fields which seemed adequate for the province having the highest crop acreage. In all, 1,450, 1,760 and 1,750 fields were sampled for wheat, rye and potatoes, respectively. Two fields of a crop were sampled per village, using the latter as the primary sample unit. Villages were sampled independently for each crop. In selecting villages, each province was stratified by counties, the number of villages selected randomly in each county being proportional to the crop acreage. A cross stratification by the size of farm was also introduced, as there was reason to believe that yields might be somewhat higher on large farms than on small farms. Sampling was, therefore, done separately for two farms size categories, the dividing line between large and small

farms being 5 or 10 hectares, depending upon the average farm size for the province. The sampling process was so controlled that the required number of fields in each farm size class would be obtained for the province as a whole. Thus in each selected village two farms were selected randomly from the one or the other category, which was predetermined. Within each selected farm, one field out of all fields growing the crop was also randomly selected.

For wheat and rye, the crop cutting plot was 1 square meter marked by a rigid square frame, open at one end, but provided with a closing bar. The frame was slipped in position to the left of the sampler, so that the rows of grain were parallel to the diagonal of the frame and the corner of the frame nearest to the sampler touched his toe. The closing bar was then placed across the open end and the ears enclosed within the frame were cut, placed in a bag and transported to a central laboratory for threshing. It is important to notice that the method of placing the frame in the crop enclosed a crop area of 1 square meter directly and not sections of rows. Consequently, no conversion factor based on average row spacing was necessary for calculating the area of the sample plot. Five such samples were cut in each selected field along the diagonal of the field, starting at a corner of the field nearest the village. A random starting point varying from two to ten paces was used for the first cutting, the remaining four cuttings being taken at uniform intervals, along the diagonal, so that its entire length was traversed. If the field was extremely large, the successive cuttings were made at an interval of 25 paces from the first cutting.

To make a proper allowance for normal harvesting and threshing losses (but not storage losses), possible immaturity of the sampled grain, since the sampling was done in advance of normal harvest, and bias due to the use of a very small cut, it was decided to get complete harvesting and threshing data for every seventh field that was sampled, in order to adjust the results from the square meter samples. The sampling teams returned to each selected field on the day fixed for its harvest by the farmer and watched the harvesting and threshing operations. The moisture content of the grain as determined from the square

meter cuttings was found to be rather high and quite variable. The yield estimates from these cuttings were, therefore, standardized in terms of clean grain, with a moisture content of 14 percent. The correction factors derived from the complete field harvesting were then applied to these yields. Actually, a common correction factor was determined for all provinces. For wheat, this was based on data from 159 fields and came to 0.88 ± 0.016. For rye, it was 0.93 ± 0.019 on the basis of 197 fields. The provincial estimates of corrected yield had standard errors ranging from 2.3 to 2.8 percent for wheat and 2.6 to 3.9 percent for rye.

The sampling design for the potato survey was similar. Unlike cereals, however, the harvest season for potato extends over a fairly long period. For this reason, the fields to be sampled were selected during the first visit to the selected village, and the sampling teams returned to this village at the time of its harvest, as intimated by the farmer, in order to take samples and also to supervise complete harvesting of fields in about 10 percent of the sampled fields. The sample plot size was a 5 meter segment of the row which was marked by cutting the row with a spade, all potatoes within this segment being lifted, rubbed free of dirt and weighed. Five such segments were sampled along a diagonal of the field, as in the case of cereals, but the starting point was taken between 5 and 10 paces instead of 2 and 10 paces. To obtain the area of the sample segments, the average row spacing in the field was estimated by determining the average width of 10 rows in the neighbourhood of the first and the last sample segments. Immediately after the samples were harvested, the whole field was harvested under supervision if it was selected for this purpose. To avoid spending a great deal of time in supervising the harvesting of extremely large fields, it was decided to restrict the operation in such fields to that part which could be harvested in two days. All potatoes were weighed in the same state in which they came from the field, but a total sample of about 50 kilograms from different parts of the field was used to determine the percentage of dirt mixed with the normally harvested crop. The ratio of the yield obtained from the complete field to the corresponding sample estimate from the same field was

used to determine the average correction factor for the latter. This was 0.88 + .012 based on the data from 185 fields. The corrected estimate of average yield per hectare determined from the sample survey had standard errors ranging from 2.2 to 3.1 percent in the different provinces. It was later decided to calculate official estimates of yield with a correction factor of 0.83 instead of 0.88, owing to the possibility that since complete field lifting was confined to relatively small fields and also because the harvesters worked under the supervision of the sampling teams, they might have picked a larger percentage of potatoes. These surveys have been continued in later years.

United Kingdom

Lastly, we may mention yield surveys on potatoes, commenced in 1948 in the United Kingdom in order to provide objective estimates of yield as a check on the official estimates (4). The survey covered most counties in England and Wales in 1948. A sample of farms was taken for the survey on a county basis. The farms were divided into three size groups on the basis of their potato acreages, a variable sampling fraction in the ratio of 1:4:8 being taken from the strata so formed. Two visits were made to all selected farms; in the first visit, two random potato growing fields were selected. At the second visit, four systematically located sample lengths of rows each approximately six feet long, were dug from each selected field and the produce weighed. A third visit was paid after lifting to a sub-sample of 1 in 10 farms and sample areas were marked out on the selected fields and carefully hand-dug to ascertain the amount of potatoes remaining in the ground. The gross yield from the samples was 10.8 tons per acre. After deducting a correction factor of 1.5 tons based on the estimate of the amount of potatoes left in the ground at normal harvest and estimation of actual area harvested in the field, the corrected yield came to 9.3 tons per acre. The corrected estimate agreed closely with the farmers' weighed yields, where the latter were available. About 1,600 fields were included in the sample survey and the sampling error of

the estimate was about + 0.12 ton per acre. The Ministry of Agriculture's estimate for the same year was 8.1 tons per acre, which was lower than the survey estimate by 1.2 tons per acre.

The brief descriptions of the sample surveys given above show that crop cutting as a method of obtaining objective and direct estimates of yield has been applied successfully under a wide range of conditions. It would thus appear that difficulties initially experienced in introducing this method in underdeveloped countries, such as ensuring a strictly random selection of plots for crop cutting and the control of the field operations with the required standard of accuracy, can be surmounted through pilot investigations and training of the field staff. Partly on account of these difficulties, estimation of crop yields by eye estimation on a large sample of fields calibrated by a limited number of carefully supervised crop cuttings has been suggested. While the approach is sound in principle, it depends for its success on the consistency and closeness of the correlation of the eye estimates with actual yield. Past experience does not show that such correlation is sufficiently stable or strong to be used in practice for an accurate estimation of yield. The adoption of sample surveys for estimating yield directly through crop cutting is therefore to be preferred under most conditions.

IV. GENERAL CONSIDERATIONS IN THE DESIGN OF CROP CUTTING SURVEYS

An efficient design for a crop cutting survey can be planned only on the background of intimate knowledge of the conditions under which a survey is to be carried out. These conditions include such aspects as the administrative set-up of the region, the type and size of the field staff available, the nature and amount of information available with the help of which the sample is to be formed, local agricultural practices, the peculiarities of the crop to be sampled and other relevant factors. The success with which crop cutting surveys can be established in practice, by securing the spontaneous co-operation of the administration and farmers alike, depends upon the skill with which sampling theory is applied to plan the survey and its field operations so as to fit in with a given set of conditions. It is obviously not possible, therefore, to lay down a single uniform design for crop cutting surveys or to frame any set rules of procedure, and it is proposed to discuss below the broad considerations that have to be borne in mind in planning these surveys.

Stratification

The aim of a crop cutting survey is generally not only to provide an accurate and unbiased estimate of the yield rate for the state as a whole, but also similar estimates for its principal administrative divisions, e.g. administrative districts as in India, crop estimating districts as in the U.S.A. or *Länder* as in Germany. These divisions thus constitute what are called the domains of study and also the principal strata. The number and distribution of crop cutting plots within these strata have to be so planned that side by side with a highly accurate estimate of yield for the entire state, estimates for individual strata may also be secured with the prescribed level of accuracy. Further stratification within these principal divisions is advantageous in increasing the precision of the yield estimates. To be effective, the yield

rate within each such stratum should be homogenous as far as possible, and the different strata should be represented in the total sample according to their sizes and the variability of the yield rate within them. The prescribed amount of sampling is then done independently within each stratum. The rate of yield is known to differ widely from place to place. A geographic stratification, therefore, suggests itself. Such stratification can be done either by sub-dividing each domain into compact arbitrary areas, or by making use of administrative sub-divisions such as counties. The latter has the advantage of convenience both in drawing the sample and in organizing the field work, and is to be preferred. Thus, in India, administrative sub-divisions with an area of 100-150 square miles (revenue circles) are commonly used as strata in crop cutting surveys, although initially larger sub-divisions of a district *(tahsils)* were used for stratification. In addition to geographical stratification, which can be recommended for general adoption in yield surveys, other types of stratification might also be considered in individual cases. Where two or more distinct varieties of a crop are grown in the region, stratification by variety might prove useful if, as is likely, yield rates of the different varieties differ considerably, and provided that accurate data for the area occupied by each variety are known. Similarly, where there is reason to believe that the yield rate of a crop is correlated with the size of farm and provided, again, that the areas occupied by the crop in different farm size categories are known, stratification by farm size would be worth trying. For efficiency of any type of stratification, the basis is that the yield rate should differ as widely as possible among the different strata, the condition for the use of such stratification being that accurate information on the size of each stratum is available as well as a list or frame of all sample units contained within the stratum from which to select the sample. The convenience of organizing the field work is always an important consideration. Without sacrificing this consideration, the possibility of securing further gains in precision by adopting a more detailed stratification of the type already adopted, or by introducing a new type of stratification, should be under continual examination in the course of sample surveys.

The next question requiring attention is that of the

method of allocation of sample units between different strata. Naturally, the most efficient distribution of the total sample among the different strata would be the one for which an estimate with desired precision is obtained at minimum cost. Where locally stationed field staff are employed for doing crop cutting work as part of their normal duties, the cost of a crop cutting survey would mainly depend upon the number of crop cutting plots to be sampled. Under this condition, the problem of optimum allocation between strata reduces to that of attaining maximum precision with a given total number of crop cutting plots. It is a well known principle that this total number should be distributed in each stratum in proportion to the product of the crop area and the standard deviation of yield in that stratum; but with strata of relatively small size, the latter would not be known with sufficient accuracy. Experience also shows that the values of standard deviation will usually not differ significantly for different strata within a district. The distribution of crop cutting plots among the strata within a district directly in proportion to the area under the crop in the strata, will thus be found to give a sufficiently good approximation to the optimum allocation. Every effort should be made to distribute the number of crop cutting plots within a district or other domain in proportion to the area under the crop in the different strata, subject, of course, to the limitations of the field staff and other facilities available. In permitting deviation from the principle of proportional allocation, it should be noted that any serious departure from this principle reduces the statistical gain from stratification appreciably. From this point of view, the overall precision of the state estimate would be unfavourably affected by the fact that a certain minimum number of crop cutting plots must be sampled within each district, or other principal division of a state for which a separate estimate of yield is desired with a given level of accuracy, as also by the fact that there is an upper limit to the number of plots that can be sampled in any such division, irrespective of its crop area, on account of the limited availability of the field staff. This is, however, inevitable.

When sampling in different strata can be done strictly in proportion to their crop areas, there is a great

simplification in the estimation of the average rate of yield over all strata as well as its precision, because these become exactly analogous to the corresponding estimates from unstratified simple random sampling. In actual practice, strict adherence to proportional allocation is not possible, and the estimates of yield rate and their precision have to be calculated separately for each stratum and then combined into corresponding overall estimates by weighting them with crop areas in the different strata.

Method of sampling within strata

For a proper random selection of crop cutting plots in each stratum, it is not practicable to make this selection in a single act of randomization and the procedure of sampling has, therefore, to be devised in successive stages. The choice of appropriate stages of sampling needs consideration and the answer has to be sought not on the grounds of precision alone, but also in accordance with organizational implications. Organizational convenience and efficiency of field work require that the sampling units should be natural units, or groups of such units, rather than arbitrary units selected out of areas formed by sub-dividing a stratum by drawing parallel, vertical and horizontal lines on the map. In countries where villages are clearly defined units with well recognized boundaries as in India, these suggest themselves as the first-stage sample units within a stratum. Fields growing the crop under study then form convenient second-stage units in which crop cutting plots may be located in random positions as third-stage units, since it is impracticable and unnecessary to harvest entire fields as sample units. In countries where the agricultural land is divided into large farms, the farm may be taken as the first-stage unit, individual fields within a farm and crop cutting plots in these fields forming the second and third stage units as before. Thus, the selection of sample plots has ordinarily to be done in three stages, as even when their selection directly from the stratum or in two stages is theoretically possible, it would involve an inordinate amount of labour in drawing the sample, and cause additional inconvenience in arranging the field work. It should be remembered, however, that with a given size of sample, the estimates have a max-

imum precision when the sample units are selected in a single stage. It is because of the practical difficulty of making such selection that selection in three stages has usually to be resorted to in crop cutting. By securing the best available frame, and subject to organizational convenience, the number of stages in which selection of crop cutting plots is made should be kept at the minimum possible.

In the actual process of selection, the requisite number of first-stage units, villages or farms, can be drawn with the help of random numbers from a complete list of these units which would ordinarily be available ready made for each administrative sub-division used as a stratum. For selecting fields, lists of fields growing the particular crop in the season will usually have to be prepared for each selected primary unit. The selection of sample units at these two stages can be done by the statistician in his office, although it is often more convenient to arrange that the selection of fields is done on the spot by the field staff. The selection of sample plots for crop cutting has always to be left to the field worker. It is necessary to lay down clear and comprehensive instructions for the selection to be made by the field staff in order to avoid any subjective bias in the selection, and to facilitate the supervisory staff's task of checking whether the instructions were strictly followed. In regard to the selection of the crop cutting plots - whatever the dimensions of the plot adopted - a theoretically sound method of locating the sample plots in random position in the selected field would be to consider the entire field as subdivided into plots of the size and shape adopted, and then to select from these, with the help of random numbers, the desired number of plots for harvesting. This method is, however, not practicable, and sample plots are usually located at points in the field which are chosen with the help of random co-ordinates taken from two adjacent sides of the field. This procedure results in over-sampling the central portion of the field at the cost of marginal areas, and if the yield rate is correlated with the relative position of the sample plot in the field, the estimate of yield obtained from plots located in this manner would be biased. In actual practice no such correlation is likely and

none has been observed in past surveys carried out in India. The above method of locating sample plots may therefore be adopted on account of its simplicity.

In selecting a sample of villages or farms and fields by simple random sampling as described above, an equal chance of selection is given to each unit irrespective of its size. It is obvious that with this method of selection smaller villages or farms and smaller fields will have a greater chance per unit area under the crop of being included in the sample than larger villages or farms and larger fields. If the yield rate of a crop is correlated with the size of a village or a farm or a field, this method of selection will lead to a biased estimate of yield, if a simple arithmetic mean is calculated from the sample. Either the method of estimation will have to be changed by giving appropriate weights corresponding to the size of the sample units, or the method of selection will have to be altered to one of selecting the sample units at each stage, with a probability proportional to a measure of size of the sample units. This process of selection requires data on the size of each sample unit in the stratum and is otherwise more complicated than simple random selection. It would not ordinarily be employed unless there was some evidence to support the belief that the yield rate was correlated with the size of sample units. This point is referred to again in a later section.

Distribution of the sample among different stages

The distribution of the sample plots between villages or farms at the first stage and fields at the second stage calls for making an effective use of the available resources. Clearly, the estimated yield for a stratum will have a maximum precision when sample plots are distributed at the rate of one per village or farm. There are, however, other considerations which counsel a reduction in the number of first-stage units and a corresponding increase in the number of second and third-stage units. The foremost of these arises from the need to ensure that the visit of the field staff to the selected primary unit is fully utilized. The limited period available for harvesting, cou-

pled with the poor communications between villages to be found in most countries, stresses the desirability of getting the maximum advantage out of a visit to the primary unit by increasing the number of plots sampled within it. This can be achieved either by selecting more fields or by locating several sample plots in each selected field. Taking the last possibility first, data from past surveys indicate that variation in yield rate between fields is far greater than that within the same field. Thus, in yield surveys on cotton in India, the mean square between fields in a village was found to be 7 to 8 times that between sample plots in the same field (15). Differences observed in several other surveys are often larger. With such differences sampling more than one plot per field is wasteful, as it adds practically nothing to the precision of the yield estimate to compensate for the additional work involved. Experience in the wheat surveys carried out in the U.S.A. described in an earlier section with the use of very small plots (1/10,000 acre) is similar (14). It was observed that the mean square between fields was 4 to 5 times as high as that between sample plots within fields. It was concluded that if only one sample plot was taken per field, it would have required sampling about 20 percent more fields, but reduced the total number of sample plots by 40 percent, in order to maintain the same degree of accuracy as was obtained with two samples per field. If one man could manage the car this would be a distinct advantage and result in a saving in the cost of the survey. Since, however, two men were required to operate the car and the crop meter, it was decided to take two samples per field which would also provide a basis for evaluating the variability within fields. This last consideration might suggest taking two samples per field where samples surveys are newly started, but a single sample per field would be found adequate for routine use. The conclusion that sampling only one plot which represents a small fraction of the area of the field is adequate also means that harvesting the entire field as a sample unit is not only unnecessary, but disadvantageous, as this would increase the cost of the field work tremendously with little gain in the accuracy of the yield estimate.

Similarly, considerations of the relative variabi-

lity between the first-stage and second-stage units, as also of the relative amount of work involved at the two stages, would determine the optimum choice as regards the number of primary units and of second-stage units within each. This point is discussed more critically in a later section, but typical results obtained under Indian conditions on the basis of relative variability among the sample units at different stages are illustrated in Table 1 which shows the number of villages required to be sampled for a given number of fields per village and plots per field for estimating the average yield with 5 percent sampling error.

Table 1. - Wheat Survey in Moradabad District (U.P.), 1944-45
(Plot size 1/92 acre)

No. of fields per village	No. of plots per field			
	1	2	3	4
	..........Number of villages..........			
1	88	74	70	68
2	63	56	54	52
3	54	50	48	47
4	50	46	45	45
5	47	45	44	43
6	46	43	43	42

It will be seen that while the number of villages decreases rather considerably as the number of fields per village is increased from 1 to 2, there is no appreciable decrease in the number of villages with the further increase in the number of fields. The increase in the number of plots selected within a field has hardly any effect on the number of villages. The results show that two to three fields per village with one plot per field is about the optimum distribution.

Size and Shape of Crop Cutting Plots

The proper choice of the size and shape of the

crop cutting plot which also determines the method of marking it, is a matter of considerable importance. It will be observed from yield surveys described earlier that the size of plots adopted falls under two broad categories. These are:

1) Small plots, circular, triangular, square or rectangular in shape and marked with portable rigid frames or other suitable apparatus. This type of plot is intended to be handled by the field investigator himself, the produce after harvest being usually bagged and sent to a central laboratory for processing. This type of plot is considered suitable where a mobile field staff has to cover a large area during the relatively short harvesting season, by rapidly moving from place to place just ahead of the normal harvest. A good example is provided by the yield surveys on wheat in the U.S.A. (13, 14).

2) Large plots, square or rectangular in shape (occasionally triangular also) and marked with the help of pegs, string, measuring tape and cross-staff. The produce is harvested and threshed in the normal manner of the farmer and with his help or with the help of hired labour. Plots of this type are used in India, Ceylon, Indonesia and elsewhere.

As the plot size increases, its variability is reduced, but the contribution to the precision of the yield estimate from this source is, relative to other sources of variability, negligible. From this point of view therefore a small plot is to be preferred; on the other hand small plots have been observed generally to overestimate yield. As mentioned earlier, the possibility of this bias is recognized in the U.S. surveys (13, 14) and special investigations suggested for studying it. In the German investigations (3), provision was made in the survey itself to control the various biases resulting from the use of small plots by harvesting a sub-sample of the selected fields completely and to provide correction factors

for yields obtained from small plots. It is interesting to note that yields estimated from small plots were 10-15 percent greater than the corresponding estimates derived from harvesting whole fields. Comparisons to test the yield estimates from plots of different sizes and shapes were made in Southern Rhodesia in 1951 and 1952 (16). They showed that small circular plots with a diameter of 5 feet 3 inches and representing an area of 1/2,000 of an acre, as also square plots with a side of 6 feet 7 inches and an area of 1/1,000 acre overestimate the yield seriously, the degree of overestimation approaching 100 percent, taking the yield obtained by harvesting the entire field as the basis.

The most extensive investigations on this subject have been carried out in India where the relative efficiency of different plot sizes was tested, first in 1944-45 in the course of a survey on wheat crop. Five different plot sizes were compared. Three equilateral triangles each of side 33 feet, 16 feet 6 inches 8 feet 3 inches respectively, marked with the help of measuring chains and pegs and two circular plots of radii 2 feet and 3 feet, marked with a specially devised apparatus consisting of a peg, a steel tape and a plumb line. The results showed that the smallest triangle and the two circular plots (all three with area less than 30 square feet) gave serious overestimates of yield. The bias towards overestimation diminished with the increase in the size of plots, but even plots of 118 square feet area were not totally free from bias. Subsequent investigations on other crops and with various sizes and shapes of plots gave similar results. The results for two investigations on paddy are reproduced in Tables 2 and 3.

Table 2. - Average yield of paddy in lb./acre for plots of different sizes
Paddy Survey in Krishna District (Madras)

Size and shape of plot	Area in sq. ft.	No. of plots	Average yield in lb./acre	Standard error of the average yield in lb./acre	Percentage overestimation
Whole field	..	108	1939.2	107.3	..
33'×13'$2\frac{1}{4}$	435.60	108	1954.1	105.0	0.8
3'(Circular)	28.29	216	2025.9	125.8	4.5
2' (Circular)	12.57	216	2113.2	129.1	9.0

Table 3. - Average yield of paddy in lb./acre for plots of different sizes
Paddy Survey in Gaya District, Bihar

Size and shape of plot	Area in sq. ft.	No. of plots	Average yield in lb./acre	Standard error	Percentage overestimation
Rectangle 33'x16$\frac{1}{2}$	544.5	206	991.54	49.37	..
Equilateral triangle of side 9.9'	42.4	412	1078.77	60.07	8.6
Isosceles right-angled triangle, equal sides 5'	12.5	412	1221.12	65.01	23.1

These show that the estimate from the large plot size (435.6 square feet = $\frac{1}{100}$ acre in Table 2) is free from bias, but plots of small size 50 square feet or less, gave an overestimation ranging up to 23 percent. Another result observed in these investigations, which has perhaps not received sufficient attention, is the effect of bias on the estimates of variability in yield between villages, fields and plots. It has been observed that like yield, estimates of components of variance due to villages and fields, which should be independent of plot size, are also inflated when plots of small size are used.

The bias observed with small plots is not their inherent property and deviations from the instructions laid down for marking and harvesting such plots can alone explain the results. Although these instructions are made as objective as possible, anyone who has had experience of measuring the length of a field and walking from a given point in the field along the direction of its length, will agree that the starting point of a plot and the direction along which it is to be laid in the field could at best be determined only approximately. Even if the same observer were to locate and mark the plot determined by a given pair of random numbers at different times in the same field, the plots may occupy different positions. The inclusion or exclusion of particular plants on the border of the plot in demarcating the plot will similarly depend upon the judgment of the field worker. The area actually cut may also vary from the one intended to be cut due to unevenly sown crops and errors in measurement. If all these deviations could be ascribed to a random element, one would expect the errors to cancel each other. The results, however, show that this is not so. They show that small plots significantly overestimate the yield, although the degree of overestimation becomes smaller with increase in size of the plot. It is obvious that the overall influence of the various factors relative to the produce harvested becomes smaller with the increase in plot size and only when the plot size is large does the bias become negligible. Whatever the reason for the bias, the use of small size plots marked by means of portable frames and other apparatus in the hands of the field staff who are normally required to do this work, is attended with considerable risk. Sometimes, small plots in the form of two or three concen-

tric circles ranging in area from 12 to 100 square feet or more, are recommended for crop cutting, as providing an internal control on the quality of the field work. It should be noted, however, that apart from the complexity involved in marking and harvesting such plots, at least the smaller circles would not only overestimate the yield, but would also give biased estimates of variability.

In view of the results quoted above, more investigations on the problem of bias arising from the use of small plots in crop cutting surveys appear highly desirable, especially under conditions where the small plot would be adopted for routine sampling on the grounds of operational suitability. In such investigations it is important that the same type of field staff as would normally be entrusted with crop cutting work should carry out the field operations connected with the location marking and harvesting of plots, after suitable training. The investigations should also be sufficiently extensive in their coverage both in regard to area and seasons. In the absence of these precautions the results might prove misleading, as it is not unlikely that in the hands of highly trained superior investigators comparisons made in limited areas might not show any evidence of biased estimation of yield from small plots.

Apart from the biases resulting from the manner of marking small plots, two further factors have to be borne in mind. The first is that the harvest and processing of the produce does not simulate these operations as carried out on the normal crop. The yield estimates, consequently, do not allow for the usual harvesting and threshing losses and may be regarded as measuring the biological yield of the crop and not its harvested yield. The second factor is related to the conditions under which small plots are used for the rapid sampling of the crop during a limited period available just before the harvest commences. Such sampling must in some cases involve either harvesting immature grain or excluding from the sample later maturing fields, as returning to these fields at the proper time for sampling would not be feasible. In either case, the estimate of yield would be biased. Again if sampling is started a little late, some of the sampled

fields might be found to have been harvested by the time the sample reaches them. The use of the large plot and the manner in which the crop is sampled with its help are free from these objections. Since the crop is harvested and processed according to the farmer's normal method, the losses associated with these operations are automatically allowed for. Secondly, the field to be sampled is selected in advance and is visited for marking and harvesting the sample when the farmer is ready to harvest it, so that there is no omission of any particular fields from the sample. For this method of crop cutting, however, it is essential to have locally stationed field staff within easy reach of the selected sample under their charge and readily available rural labour to help in the crop cutting operations. Where such conditions do not exist, and the crop must be sampled a little ahead of the normal harvesting period through a rapidly moving staff using small plots, a necessary precaution is to arrange for complete harvesting of the field under supervision in a subsample of the selected fields, in the manner used in yield surveys in Germany (3), in order to adjust the results based on small plots. Instead of whole fields, plots large enough to be handled by normal harvesting methods will also provide a basis for adjustment. Without such adjustment, yield estimates derived from surveys involving small plots would generally be biased.

Adjustment for Yields

So far, we have considered field crops for which the appropriate sampling unit is a plot of suitable dimensions. For certain crops which are grown in widely spaced rows such as potatoes or sugar cane, a section of a single row or of two or more adjacent rows would be a more suitable unit. The total row length harvested will have to be multiplied by the average spacing between rows as estimated for the sampled field in order to work out the yield of the sample on an area basis. No work seems to have been done so far on the optimum size of sample units for such crops and it would appear that row lengths which give an area equivalent to that of a large plot, (say, 1/100 or 1/50 acre) may be regarded tentatively as being

suitable. Fruit trees and other tree crops would also require special treatment as far as the choice of a sample unit is concerned. Here, an individual tree or a cluster of adjoining trees would seem to form a suitable sampling unit and the estimate of yield may be calculated on a per tree basis rather than on an area basis, villages or farms being the first stage unit as in the surveys with field crops, and orchards or fruit gardens being the second stage units.

An adjustment to the yield estimate necessary in all surveys on grain crops is the allowance for the moisture content of the grain. Since the grain on the harvesting day contains moisture, and since it is neither safe nor practicable to store the harvested material for drying before threshing it, a sample of the grain threshed from the harvested produce on the same day it is harvested should be stored and re-weighed after drying, so that an estimate of the yield of grain can be calculated with a standard moisture content. It must be emphasized, however, that after the requisite adjustments, the yield estimate from crop cutting provides an unbiased estimate of the produce at harvest. Losses occur in the cartage and storage of crops and in some cases these are quite large, so that the estimated yield is an overestimate of the total crop which ultimately reaches the customer. Allowance for losses after harvest is a separate problem, but it should not prove insuperable.

An important adjustment in the case of the rice crop is that for converting the yield of paddy grain (rice in husk) into the yield of dehusked rice. Dehusking is done in rice mills as well as by hand pounding, the losses in the two cases differing appreciably. In rice mills the percentage of clean rice to paddy grain is usually not more than 65 percent and may go down to 55 percent or lower. With hand pounding the loss is less than in milling, the proportion of clean rice being 70 percent or more. On account of such wide fluctuations in the conversion factor according to the method of milling and also because of varietal and regional differences in this factor, its careful estimation by sampling under mill and hand pounding conditions and weighing by the relative proportions of paddy grain treated by the two methods is important. The investigations may also be extended to different varieties in a region if the proportion

of the total produce from each variety is known. These studies should proceed side by side with yield surveys on paddy in order to enable the expression of the production estimates in terms of clean rice without bias.

Mixed crops

Mixed crops are of common occurrence, especially in underdeveloped countries. They constitute an important factor in estimating yield by crop cutting but offer no serious difficulty. Crop mixtures tend to run into standard patterns, as for instance, wheat-barley and wheat-yam mixtures in India or the maize-yam-cassava mixture in parts of Africa. It is obvious that the entire area in which the crop under study is being grown, whether pure or in a mixture with other crops, must be sampled for crop cutting. If the proportion of each category is appreciable and is known, or can be estimated, it would be an advantage to stratify the total crop area into the two categories and sample separately within each. Whether this is done or not, the size of the crop cutting plot fixed should refer to the gross area under the crop and not the net component of this area occupied by the crop being sampled. Determination of such components is usually arbitrary and frequently impossible. The yield of the crop gathered from the crop cutting plot will then be credited to the gross area of the plot and would also be multiplied by the gross area under the crop in calculating the total production of the crop.

Sometimes trees like coconut, palm oil or other fruit trees may be scattered among other field crops, or yams or cassava plants may be grown in distinct hills in fields of maize. An alternative which appears feasible under these conditions and is worth exploring in studying the yield of such crops is to treat the individual tree or hill as a sample unit and estimate the yield per tree or per hill. For estimating the total production of the crop, however, it will be necessary to estimate the total number of trees or hills in the area under survey.

V. SOME TECHNICAL CONSIDERATIONS

The more important aspects of the planning of crop cutting surveys have been discussed in broad outline in section IV. Some technical considerations involved in the analysis of survey data are now given as being particularly relevant to the planning of these surveys.

It will be clear from the earlier sections that a suitable design for crop cutting surveys would be what is called stratified multi-stage sampling. Each region or domain for which an estimate of yield rate is required, which would ordinarily be a principal administrative division of the state, would be stratified according to its administrative sub-divisions and within each sub-division a number of primary sampling units, viz. villages or farms, would be selected at random. For the purpose of the present discussion villages will be referred to as the primary units. This number would be approximately proportional to the crop area in the sub-division. Within the selected primary unit a small fixed number of second-stage units, namely the crop growing fields, would be selected next and within each selected field usually one crop cutting plot of prescribed size and shape would be marked in a random position and harvested, although more than one such plot might be selected per field in the initial stages of the survey.

Estimation of mean yield

Where there is evidence to suggest that the yield rate is correlated with the size of the sample units, both the first-stage and the second-stage units may be selected with probability proportional to their crop areas. Within a selected field, the crop cutting plot may be located in a random position. With one plot per field, the method of estimating the yield rate is to compute the simple arithmetic mean of plot yields in the sample within each stratum, thus

$$\bar{y}_s = \frac{\sum^{n} \sum^{m} y_{ij}}{nm}$$

where y_{ij} denotes the plot yield from the j^{th} field of the i^{th} village, m denotes the number of fields selected for sample harvesting in a village, and n denotes the number of villages in the sample. Weighted by the crop areas in the different strata, it provides the estimate of the yield rate for the region. The estimate is unbiased for the system of sampling with probability proportional to the size of sample units.

The variance of the mean yield is given by

$$V(\bar{y}_s) = \frac{\sigma^2_b}{n} + \frac{\sigma^2_w}{nm}$$

Where σ^2_b and σ^2_w stand for true variances between village means per plot and between field means per plot within villages. This variance can be estimated from the sample by

$$\text{Est } V(\bar{y}_s) = \frac{s^2_b}{nm}$$

Where s^2_b is the mean square between villages on per plot basis in the analysis of variance of the sample.

When villages and fields are selected with equal probability by simple random sampling, as would generally be the case, the simple arithmetic mean of plot yields does not give a strictly unbiased estimate of the stratum value of the yield rate, since, as stated previously, this process of selection gives smaller villages and smaller fields relatively a larger chance of selection per unit area under the crop. The simple arithmetic mean would in this situation provide an acceptable estimate of the yield rate only if it is known that there is no correlation between the size of villages and of fields on one hand and yield rate on the other. Indian experience shows that the correlation is absent, but the position will need to be examined in individual circumstances. The justification for the use of the simple arithmetic mean is the extreme simplicity of its computation and its high efficiency.

In considering the variance of the simple arithmetic mean from a sample selected with equal probabil-

ity, it may be noted that the value of m will usually be small compared with the total number of fields under the crop in a village. Likewise, the area actually harvested will be negligible compared to the area under the crop in a field. The variance of the estimate may consequently be written as

$$V(\bar{y}_s) = \left(\frac{1}{n} - \frac{1}{N}\right) S^2{}_b + \frac{1}{nm}\left(S^2{}_f + \frac{1}{p} S^2{}_p\right)$$

Where $S^2{}_b$, $S^2{}_f$ and $S^2{}_p$ are the population mean squares between true village means per plot, true field means per plot and true plot yields respectively, and N is the total number of crop growing villages in the stratum. With only one plot per field, we may substitute $S^2{}_w$ for $S^2{}_f$ + $S^2{}_p$ in the above formula. The estimate of the various population mean squares may be obtained from the analysis of variance of plot yields. For a sample of n villages, m fields per village and p plots per field, the analysis of variance of the nmp plot yields takes the following form:

Variation due to	*Degrees of freedom*	*Mean square*	*Estimate of*
Villages	n-1	$s^2{}_b$	$mpS^2{}_b + pS^2{}_f + S^2{}_p$
Fields within villages	n(m-1)	$s^2{}_f$	$pS^2{}_f + S^2{}_p$
Plots within fields	nm(p-1)	$s^2{}_p$	$S^2{}_p$
Total	nmp-1		

When p = 1, i.e. with one plot per field, an estimate of the variance between plots, $s^2{}_p$, is not available as this component is included in the estimate of variance between fields and is inseparable from the latter. The latter may be termed $s^2{}_w$ and is the estimate of the population var-

iance S^2_w between fields; the mean square s^2_b is then the estimate of $mS^2_b + S^2_w$. The variance of the estimated mean is thus calculated from

$$\text{Est } V(\bar{y}_s) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{s^2_b}{m} + \frac{s^2_w}{Nm}$$

When the sampling fraction at the first-stage is also small, so that $\frac{n}{N}$ is negligible, the variance is given by $\frac{S^2_b}{n} + \frac{S^2_w}{nm}$ and its estimate by $\frac{s^2_b}{nm}$, which is directly calculable from the analysis of variance of the sample.

If under the method of sampling with equal probability the simple arithmetic mean is found to be biased, an unbiased estimate can be obtained by calculating the weighted mean given by

$$\bar{y}'_s = \frac{1}{A}\cdot\frac{N}{n}\sum^{n}\frac{M_i}{m}\sum^{m} A_{ij}y_{ij}$$

where, A is the total area under the crop in the stratum, A_{ij} is the area of the j^{th} field in the i^{th} village and M_i is the number of fields under the crop in the i^{th} village.

The variance of the weighted mean is estimated approximately by

$$\text{Est } V(\bar{y}'_s) = \frac{s'^2_b}{n}$$

where s'^2_b is the weightedmean square between village mean yields per plot and is given by

$$s'^2_b = \frac{1}{n-1}\sum^{n}\left(\frac{A_i\,\bar{y}'_i - \frac{\sum^{n} A_i\,\bar{y}'_i}{n}}{\bar{A}}\right)^2$$

where $\bar{y}'_i$ is the weighted mean estimate for the i^{th} village, A_i is the crop area in the i^{th} village and $\bar{A}$ the average crop area per village. The variance of the weighted mean is larger than that of the simple arithmetic mean. Percentage standard errors calculated for the two types of mean

from a few yield surveys carried out in India are included in Table 4.

Table 4. - Percentage standard errors of different estimates of mean yield

	Simple arithmetic mean $\bar{y}_s$	Weighted mean $\bar{y}'_s$	Ratio estimate $\bar{y}''_s$
Wheat (U.P.), 1947-48	3.7	14.0	4.7
Wheat (Delhi), 1948-49	2.5	10.0	5.7
Cotton (Madhya-Pradesh),			
1944-45	5.5	15.0	11.3
1945-46	6.9	14.0	13.2

These results show that the simple arithmetic mean is far more accurate than the weighted mean calculated from the same sample and is to be preferred provided the bias in the former can be neglected. The nature and magnitude of this bias is illustrated in Table 5 from a large scale Indian survey.

Table 5. - Mean yield in wheat survey in Punjab, 1943-44

District		Yield in lb. per acre	
		Simple mean $\bar{y}_s$	Weighted mean $\bar{y}'_s$
1	Amritsar	1,029	1,041
2	Gurdaspur	829	862
3	Jullundur	839	881
4	Hoshiarpur	804	796
5	Ludhiana	1,247	1,246
6	Ferozepur	1,052	1,079
7	Ambala	854	820
8	Karnal	839	868
9	Hissar	1,090	1,142
10	Rohtak	1,001	997
11	Gurgaon	766	752
	State	920	927

This table gives a comparison of the yield estimate in large samples, based on the simple arithmetic mean and the weighted mean. The standard error of the difference between the two estimates is known to be of the order of 6 to 8 percent; the table shows that not only do the differences change sign from district to district, but also their magnitude is negligible compared to their standard errors. The table is typical of the results usually obtained and shows that bias arising from the use of the simple arithmetic mean is negligible for all practical purposes.

Where on account of its bias the simple arithmetic mean does not provide a good estimate of the yield rate, an alternative to the weighted mean $\bar{y}'_s$, which we have seen is far less efficient, is the ratio estimate given by

$$\bar{y}''_s = \frac{\sum^{n} A_i \dfrac{\sum^{m} A_{ij} Y_{ij}}{\sum^{m} A_{ij}}}{\sum^{n} A_i}$$

The estimate is not strictly unbiased, but the bias diminishes rapidly as the size of the sample is increased, so that in reasonably large sample surveys the bias would be negligible. Its chief advantage lies in its accuracy, which is distinctly greater than that of the weighted mean and often approaches that of the simple arithmetic mean. This will be clear from the percentage standard errors for the ratio estimates which are shown side by side with those of the other two estimates in Table 4.

The variance of the ratio estimates can be estimated from the sample approximately by

$$\frac{s''^2_b}{n}$$

Where s''^2_b is the mean square between village mean yields per plot calculated as ratio estimates as shown in the formula for the ratio estimate $\bar{y}''_s$, and is given by

$$s''^2_b = \frac{1}{n-1} \sum^{n} \frac{A_i^2}{\bar{A}^2} (\bar{y}''_i - \bar{y}''_s)^2$$

Efficiency of Stratification

For studying the efficiency of stratification with the help of data from a stratified sample, the analysis of variance of plot yields similar to the one given above pooled over the different strata is a useful method, provided the number of villages are distributed among the strata in proportion to the crop areas and the order of magnitude of variation in the yield rate within the different strata is the same. If a total of n sample villages are allocated at the rate of n_1, n_2, ... n_k to k strata, the pooled analysis of variance for the nmp plot yields takes the following form:

Variation due to	*Degrees of freedom*	*Mean square*	*Estimate of*
Strata	$k - 1$		
Villages within strata	$k\Sigma(n_i - 1)$	s^2_b	$mpS^2_b + pS^2_f + S^2_p$
Fields within villages	$k\Sigma n_i(m - 1)$	s^2_f	$pS^2_f + S^2_p$
Plots within fields	$k\Sigma n_i m(p - 1)$	s^2_p	S^2_p

n_i being the number of villages in the sample from the i^{th} stratum. With only one plot per field a single mean square s^2_w represents the last two items; s^2_b and s^2_w are then estimates of $m\,S^2_b + S^2_w$ and S^2_w respectively.

The estimated variance of mean yield per plot from the stratified sample is

$$\frac{s^2_b}{nmp}$$

assuming the sampling fraction to be small at all stages.

If the sample of n villages were to be selected randomly without any stratification, the component of variation due to strata in the above analysis of variance would have formed part of the error of the estimate and the variance of the mean yield would then be calculated from the mean square obtained by adding together the sum of squares for the strata and villages in the above analysis, and dividing them by the appropriate number of degrees of freedom, viz. $k-1 + k\Sigma(n_i-1)$ which equals $n-1$. Since the mean square between strata is usually larger than between villages within strata, the new mean square on which the variance of the unstratified sample is to be based, would also be larger than s^2_b. The ratio of these two mean squares will thus indicate the relative efficiency of the stratification adopted.

The use of the analysis of variance is illustrated with the results of a cotton yield survey carried out in Beras division of Madhya Pradesh in India in 1947-48. The analysis of variance of plot yields was as follows:

Table 6. - Analysis of variance of cotton yield, lb. per plot

Source of variation	Degrees of freedom	Mean square
Tahsils within districts	17	4,313
Revenue circles within *tahsils*	68	2,254
Villages within revenue circles	157	1,846
Fields within villages	492	1,134

Villages and fields formed the sample units at the two stages, only one plot of 1/10 acre size being located in each field for harvesting. Administrative districts were domains of study and were stratified by revenue circles which are sub-divisions of *tahsils*. Stratification by *tahsils* which are the larger sub-divisions of a district was also possible. The error mean square is seen to be 1,846. If there was no stratification within a

district, the error mean square would have been obtained by pooling together the sums of squares for the first three items in the table and dividing by the corresponding degrees of freedom viz. 242. The value of this mean square would be 2,134. If *tahsils* had been used as strata instead of revenue circles, the error mean square would have been obtained by pooling the sums of squares for the second and third items and its value would be 1,969. Thus without any stratification within the districts the error mean square would have increased by 16 percent and if the stratification had been confined to *tahsils* it would have increased by 7 percent.

Where the allocation of the sample villages to the different strata is far from proportional and also where there are distinct differences in the variability of the yield rate among the different strata, the analysis of variance approach is not useful. We have then to set down expressions for weighted variances of sample means for the stratified sample as drawn and for a corresponding hypothetical unstratified sample.

The estimate of the first variance may be written as

$$\sum^{k} W_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) \hat{S}_{hb}^2 + \sum^{k} W_h^2 \frac{\hat{S}^2_{hw}}{n_h m}$$

where W_h is the proportion of the total area under the crop in the h^{th} stratum,

$\hat{S}^2_{hb}$ is the estimate of the population mean square for village mean yields per plot in the h^{th} stratum

and $\hat{S}^2_{hw}$ is the estimate of the population mean square for plot yields within villages in the h^{th} stratum.

The variance of the mean of an unstratified sample is given by

$$\left(\frac{1}{n} - \frac{1}{N}\right) \hat{S}_b'^2 + \frac{\hat{S}_w'^2}{nm}$$

Where

$$\hat{S}_b'^2 = \sum^{k} \left(W_h - \frac{W_h}{n_h} + \frac{W^2_h}{n_h} - \frac{W^2_h}{N_h}\right)\hat{S}^2_{hb} - \sum^{k} \frac{(W_h - W^2_h)}{n_h m} \hat{S}^2_{hw} + \sum^{k} W_h \bar{y}_h^2 - \bar{y}_s^2$$

$\hat{S}'^2_w$ = estimate of the population mean square between plot yields within villages over all strata and $\bar{y}_h$ and $\bar{y}_s$ are the mean yield per plot within h^{th} stratum and the weighted mean yield over all strata respectively.

A ratio of the two variances then gives the gain in precision resulting from stratification.

Allocation of Sample Units at Different Stages

Another important question in planning a crop cutting survey is that of allocation of the crop cutting plots among the different stages of sampling. True variances for villages, fields and plots estimated from the pooled analysis of variance of a sample shown above can be substituted with the help of the relationships indicated there in the following formula for the variance of mean yield:

$$V(\bar{y}_s) = \frac{S^2_b}{n} + \frac{S^2_f}{nm} + \frac{S^2_p}{nmp}$$

By using this formula, the variance of the mean yield to be expected from different values of n, m and p can be calculated. Alternately, the values of n, m and p required for attaining the requisite magnitude of the variance of the sample estimate can be determined. The latter procedure was employed for calculating the number of villages required to be sampled for given numbers of fields per village and plots per field for estimating the mean yield with 5 percent standard error shown in Table 1. The total number of villages, n, thus calculated should be allocated to the different strata in proportion to their crop areas.

Consideration of the order of magnitude of the sampling error with which it is desired to estimate the mean yield has to be supplemented by considerations of cost in order to obtain guidance on the optimum number of sample plots and its distribution between villages and fields. Broadly, the cost of the survey will include expenditure on salaries and travel of the field staff and labour required for field operations connected with the crop cutting. As a first approximation, we may assume the first component to

be proportional to the number of villages selected, and the second to the number of crop cutting plots sampled within each selected village. The cost of the survey may thus be represented as

$$C = c_1 n + c_2 nm$$

Where c_1 and c_2 are known constants determining the rates of cost at the two stages. The variance of the sample mean when one plot is located in a field is, as we have seen earlier, given by

$$V(\bar{y}_s) = \frac{S_b^2}{n} + \frac{S_w^2}{nm}$$

The optimum value of m can then be obtained by minimizing the cost plus variance function $V + \lambda C$. The value of n can be calculated from the equation for cost, if this is given, or from the equation for variance if the latter is prescribed. The optimum values of m and of n for given variance are shown below:

$$m = \sqrt{\frac{c_1}{c_2}} \cdot \frac{S_w}{S_b}$$

$$n = \frac{S_b^2 + \frac{S_w^2}{m}}{V}$$

where the values S_b^2 and S_w^2 have to be determined from past data and V is the value of variance prescribed for the sample mean. The cost of the survey can further be calculated by substituting the values of n and m in the expression for cost and this cost will be minimum for attaining the desired level of accuracy in the estimate.

Experience in India shows that owing to poor communications between villages, it is not practicable to visit more than one village during the day. For the same reason, it becomes important that the field worker should use his time fully once he has reached a village. Consequently, where the field staff is drawn from the administrative machinery of the government, with each man responsible for work within his own jurisdiction on a fixed

monthly salary and travel allowance, the cost of travelling time and transport to a selected village becomes equivalent to the cost of daily salary and travelling allowance. The value of c_2 is clearly the value of labour of sample harvesting a plot. On the basis of salaries and allowances paid to the field staff, and of the rates of daily wages for rural labour in India, we may take the values of c_1 and c_2 in the ratio of 7:2 and from past surveys the variance between fields as being about $1\frac{1}{2}$ times as much as the variance between villages. We then find that the optimum value of the number of fields per village is, to the nearest integer, 3. It is interesting to note that this value is about the same as that derived from a consideration of Table 1.

Depending upon the structure of the survey, communications and other conditions of the surveyed regions and the type of field staff employed, appropriate expressions for cost can be set up and the necessary variance components determined from the data of pilot field investigations. The data can be analysed in order to determine the optimum number and distribution of sample units at the different stages of the survey design. Such analysis is very useful as forming a quantitative basis for ensuring most efficient and economical expenditure of resources on crop cutting surveys.

The technical considerations discussed above are intended merely to indicate how modern sampling theory can be applied to the planning of crop cutting surveys and the analysis of their results. Excellent publications giving comprehensive accounts of this theory and its varied applications are now available.

VI. FIELD WORK AND PERSONNEL

The work of planning and executing crop cutting surveys and the analysis of their results is generally shared by the statistical and field staff entrusted with this work. Decisions regarding the design of the survey, including the kind and number of strata, the type, number and distribution of sample units at different stages and also the actual selection of the first and second-stage units, namely the villages or farms and fields, are normally the responsibility of the statistician. He may prepare, in addition, suitable instructions for field work, train the field staff and also participate in the supervision of field work during its progress. The work of the field staff includes conducting various crop cutting operations, such as locating and marking crop cutting plots in the selected fields and harvesting and processing their produce and reporting the results in forms prescribed for this purpose. Sometimes, selection of fields to be sampled is also carried out by the field staff concerned. This may be done in a preliminary visit to the selected primary unit by listing with the help of a cadastral map of the fields, if one is available, or by preparing a sketch map, the fields growing the crop to be sampled. The requisite number of fields are then selected out of this list with the help of random numbers. If the primary unit is large, and complete listing of the fields under the crop is not feasible, the selection is based on information on the total number of fields included in the primary unit irrespective of the crop. The requisite number of random numbers are then drawn within this total number, and if the fields corresponding to the numbers drawn are not found to grow the particular crop on inspection, successive fields in serial order are examined until fields which grow the crop are located. A second visit will be made when the fields are ready for harvest and the crop will be sampled by marking crop cutting plots in the fields. Where a mobile field staff is rapidly sampling the crop just ahead of the harvest, selection of fields and sampling of the crop are more conveniently done in the course

of a single visit to the primary unit.

Some uniform definition of a field is necessary in order that no bias may creep into the selection. A field for the purpose of crop cutting may be taken to mean a distinct patch or portion of land which has no bunds within it other than small irrigation bunds and which is demarcated on all its sides either by means of a bund or narrow strip of grass or uncultivated land or by means of a crop or crops different from the one grown in the patch. Unambiguous instructions are also necessary for the guidance of the field staff for locating sample plots in the selected fields, marking them in the prescribed manner for harvesting and threshing, drying and weighing the produce. The manner of conducting these operations naturally depends upon the conditions of the survey and the local agricultural practices. Instructions prepared for yield surveys on wheat in India are reproduced below, with slight modifications, as an illustration.

Location and Marking of Crop Cutting Plot

In each selected field one rectangular plot of $33' \times 16\frac{1}{2}'$ is to be located at random. This is not to be done earlier than on the day already fixed for harvesting. Before a plot is located, make sure that the field is the one already selected by you. The procedure of locating a random plot is as follows:-

Stand facing north with the field in front of you and to your right. This corner of the field (south-west corner) will be the starting point. When the sides of a field are not north-south and east-west, the direction towards which you should face to determine the starting point will be approximately north. For convenience fix a peg at the starting point.

Beginning from the starting point measure by means of a tape, the length and the breadth of the field in feet, fractions being rounded off

to the next higher number. Next deduct 33 from length and 17 (in place of $16\frac{1}{2}$ for convenience) from the breadth and obtain respective remainders. Now select two random numbers, one each for the length and breadth and not exceeding the respective remainders obtained by consulting the random number lists of appropriate digits. The pair of random numbers so selected will determine the south-west corner of the plot to be marked. The random numbers should be ticked off as they are used. A circle should be marked around the selected numbers.

Suppose the pair of random numbers selected for locating the plot is say (064,23). To mark the plot by means of this pair of random numbers walk from the starting point of the field along its length and stop at a distance of 64 ft. Having arrived at this point, walk into the field along the breadth and stop at a distance of 23 ft. Fix a tall straight bamboo peg at this point, which will be the starting point for the plot.

Tie a string to the peg fixed at this point, stretch it along the direction of the length of the field away from its starting point and measure 33 ft. along it and fix another peg at this point. Replace this peg by the cross-staff. Turn the string round the cross-staff, stretch it at right angles away from the starting point of the field and measure $16\frac{1}{2}$' along it. This will be the third corner of the plot provided the distance of the diagonal over the ear-heads. Fix a third peg at this point, use the cross-staff, turn the string round the peg, and stretch it parallel to the direction of the length of the field towards the starting point measuring 33' reaching the fourth corner of the plot. This will be the point for the fourth peg provided the diagonal distance from the second peg to this point is exactly 36' 10''. Check the diagonal. Fix a fourth peg at this point, turn the string round the peg, stretch it to the starting corner of the plot and tie it to the first peg. Check

the distance between the fourth and the first peg which should be $16\frac{1}{2}$'. The pegs should be tall, straight and firmly fixed into the ground. See that the string is fully stretched on all the four sides. Carefully lower the string to the level of the ground.

If the whole of the plot does not fall within the field owing to slight irregularities in its shape, reject the pair of random numbers both for the length and breadth and mark the plot over again using a fresh pair of random numbers obtained by reading further down the columns. If, however, a located plot falls wholly within the field, take the position of the plot so marked as final notwithstanding whether the crop inside the plot is poor or otherwise.

Harvesting and Other Operations

Harvest the crop which is only within the boundary of a plot. Any bunches of tillers which lie on the boundary of the plot, partly inside and partly outside, should be carefully separated. It is advisable not to allow the surrounding crop of the field to be harvested and removed to the threshing ground.

Complete the harvesting before noon, collect all the harvested produce without leaving any ear-heads in the plot and spread it on a piece of cloth for a few hours before threshing and winnowing it. The produce should be threshed by trampling under feet or by beating with a wooden rod and winnowed with flat winnowers. Take care to see that there is no loss in the produce at the various stages, viz. harvesting, separating,carrying from the field to the threshing ground, threshing, winnowing, cleaning and weighing. Particular care should be taken to see that all grain is separated from the ear-heads as also obtained free

from dust. Weigh the cleaned produce carefully. Weighing should be to the nearest half *Chhattank* (an Indian weight approximately equal to an ounce). In the case of mixed crop, the weight of component crops should be recorded separately. Complete all the operations for a field on the same day but where the produce is moist and it is difficult to separate the grain from ear-heads, it should be allowed to dry up for a day or two under the care of the village headman.

Record the results of weighings obtained on the day of threshing in the form provided for the purpose and despatch the form immediately as per instructions contained therein.

Allowance for Drying

Operations for estimating the allowance for drying will be conducted for the first plot harvested in each of the first two villages allotted to you. The operations are as follows:-

Immediately after the produce of a plot is weighted, store a sample of exactly one *seer* (approximately equal to 2 lb.). If the total is less than this quantity store the whole of it. Seal the bag, label it properly and deposit it in your office. The bag should be exposed to the sun every day unopened for about a fortnight till the inside grain is well dried.

At the end of the period, weigh the contents of the bag carefully, record the results of re-weighing and other particulars in the form provided and despatch it as instructed therein. Take care to see that the re-weighing is done with the same set of scales and weights as were used at the time of initial weighing. Return the produce to the owner after the grain has been re-weighed.

For success in crop cutting work, it is essential that the statistical staff concerned with the planning of the surveys, and the field staff entrusted with their execution, be familiar with the agricultural conditions of the region. It is a great advantage if the latter, in addition, are known to the farming community of the area, speak their language and are normally residents of the area in which they work. Experience shows that for this reason administrative officials of agricultural, land records or other similar departments located in rural areas suit admirably for this work done as part of their normal duties. The field staff engaged in crop cutting, as also that entrusted with its supervision, require to be given intensive training in the method of work, such training being invariably accompanied by spot demonstrations. It is useful to associate with the training senior officers of the departments interested in this work, since it helps them to keep a vigilant eye on the work of the staff and to make fruitful suggestions for improving its efficiency. The training should preferably form a short refresher course at the beginning of each season, when the relative significance of different mistakes in the conduct of crop cutting can be stressed, as for instance by explaining how a slight error in the dimensional measurement of a selected plot is undoubtedly more harmful than a smaller error in the measurements relating to the length and breadth of the field containing the plot. Instances of conscious or unconscious departure from instructions laid down and its possible consequences can be brought home to the field staff, and strict compliance with these instructions insisted upon.

Each member of the staff requires certain equipment for crop cutting. This ordinarily consists of tables of random numbers, a measuring tape, string, pegs and cross-staff for marking the plot if it is large, or a suitable frame or other apparatus with which the plot is to be marked, a piece of cloth on which to spread the produce, scales with a standard set of weights for weighing it and cloth bags for storing a sample of the produce to estimate its moisture content. For small plots where the produce is to be bagged and sent to a central laboratory, the equipment includes a pair of scissors for clipping the plants and strong bags in which to dispatch the harvested material.

Provision of proper equipment is an important factor in avoiding slipshod field work.

For successful introduction of the crop cutting method in a new area, it is of utmost importance to secure the co-operation of the farmers, village headmen and other interested persons and to build up enthusiasm for this work in the entire rural community. The farmers' co-operation is essential where crop cutting plots are large and are harvested at the time of normal harvest with their help. Even where small plots are sampled and the field worker manages these himself, he would not be able to do this without their goodwill. Wide publicity among the farmers on the objects of the survey, the details of the procedures adopted, and the advantages that the farmers themselves are likely to secure by timely and accurate information on current crop yields is helpful in securing this co-operation. A strict practice of returning the entire produce of the sample plot to the owner is found to be a factor in gaining their confidence. Experience shows that although farmers in a backward area may be initially suspicious of such surveys, they offer their co-operation when the right approach is made.

To secure reliable primary data from crop cutting surveys, careful supervision of field work is essential. In addition to the normal supervisory staff, the statistical staff connected with the surveys may also participate in the supervision. The need for associating all supervising staff with the training program has already been mentioned. Further, instructions for field work and forms for recording information and data can be so prepared that the work of checking by the supervisory staff is facilitated. For instance, if the field worker is required to record the particular column of random numbers which he uses for locating the sample plot in the selected field, it would be easy for the supervisor to find out by spot inspection whether the plot was correctly located. To make it more effective, it is possible to organize the supervision itself on a random sampling basis, so that the crop cutting plots sampled under supervision form a representative sub-sample of all plots. This serves as a statistical check on the quality of the field work and provides an adjustment for the results, if necessary.

VII. CONCLUSION

In its aim of building up a reliable annual picture of world agricultural production, FAO has already recognized the vital role of objective methods of crop estimation and is making efforts in various directions to promote their use. International training centers on statistical sampling in agriculture accompanied by illustrative sample surveys and other field demonstrations have been held in Ecuador, Thailand and Nigeria and similar centers have been planned for 1954. A seminar on advanced sampling was arranged in Sweden, in which crop cutting methods of estimating crop yield formed an important item. Individual countries that are prepared to take up improvement of their agricultural production statistics are provided with expert assistance for organizing sample surveys. A country-wide sample survey for estimating the area and yield of rice has already been carried out in Ceylon and the work is being continued. A sample survey with similar objectives but on a small scale was concluded successfully in Thailand and has been referred to earlier. A sample survey for estimating rice production is in progress in Indonesia and plans for sample surveys for crop estimation are being made for Iran and Sweden. A vigorous continuation and expansion of these activities over the next few years would contribute greatly to achieving FAO's aim of collecting accurate statistics of world crop production and what is perhaps more important in helping countries to place their production data on a sound footing.

VIII. REFERENCES

(1) U.S. BUREAU OF AGRICULTURAL ECONOMICS AND U.S. PRODUCTION AND MARKETING ADMINISTRATION
The Agricultural Estimating and Reporting Services of the United States Department of Agriculture.
Washington, U.S. Department of Agriculture, 1949. (U.S. Department of Agriculture Miscellaneous Publication No. 703).

(2) STATISTISK TIDSKRIFT
A New Method of Estimating Crop Yields. Stockholm, 1952. (Extract from Statistisk Tidskrift No. 2).

(3) HENDRICKS, WALTER A.
Objective Yield Estimates for Winter Wheat, Winter Rye and Potatoes in the UK/US Zone of Germany.
Frankfurt, Working Paper of the Bureau of Agricultural Economics, U.S. Department of Agriculture, 1948.

(4) YATES, F.
Agriculture, Sampling and Operational Research.
Berne, 26th Session of the International Statistical Institute, 1949.

(5) THE CENTRAL AFRICAN STATISTICAL OFFICE
Report on the Sample Census of African Agriculture of Southern Rhodesia. Salisbury, 1951.

(6) INDIAN COUNCIL OF AGRICULTURAL RESEARCH
Sample Surveys for the Estimation of Yield of Food Crops. New Delhi, 1951. (Bulletin No. 72).

(7) SUKHATME, P.V.
Sampling Theory of Surveys with Applications.
New Delhi, Indian Society of Agricultural Statistics; and Ames, Iowa (USA), Iowa State College Press, 1953.

(8) INDIAN STATISTICAL INSTITUTE
Report on Bengal Crop Survey 1944/45. Alipore, West Bengal Government Press, 1948.

(9) MAHALANOBIS, P.C.
Recent Experiments in Statistical Sampling in the Indian Statistical Institute. London, Journal of the Royal Statistical Society, 1946. (Vol.CIX, Part IV, pp.325-370).

(10) KOSHAL, R.S.
Report on the Sample Survey of Paddy Yields and Acreages in Ceylon. Rome, Food and Agriculture Organization, 1954. (In preparation).

(11) FOOD AND AGRICULTURE ORGANIZATION
Report on Far Eastern Demonstration Center for Agricultural Statistical Sampling. Rome, 1953. (ETAP Report No. 137).

(12) DOUGLAS, A.J.A. AND TENNANT, R.K.
Basutoland Agricultural Survey, 1949/50. Maseru, Basutoland Government, 1952.

(13) KING, ARNOLD J. AND JEBE, EMIL H.
An Experiment in Pre-Harvest Sampling of Wheat Fields. Iowa (USA), Agricultural Experiment Station, 1940. (Research Bulletin No. 273).

(14) KING, ARNOLD J., McCARTY, DALE E. AND McPEEK, MILES
An Objective Method of Sampling Wheat Fields to Estimate Production and Quality of Wheat. Washington, U.S. Department of Agriculture, 1942. (Technical Bulletin No. 814).

(15) SUKHATME, P.V. AND PANSE, V.G.
Crop Surveys in India - II.
Journal of the Indian Society of Agricultural Statistics, 1951. (Vol.III, No. 2, pp. 98-168).

(16) SHAUL, J.R.H.
Lectures on Planning African Agricultural Sampling Surveys. Ibadan, Nigeria, African Training Center for Agricultural Statistics. Rome, Food and Agriculture Organization, 1953.